HYBRID-1D MACROBLOCK PREDICTION FOR VIDEO COMPRESSION

Jean-Marc Thiesse, Joël Jung, Marc Antonini

HYBRID-1D MACROBLOCK PREDICTION FOR VIDEO COMPRESSION

J-M. Thiesse[1], *J. Jung*[1], *and M. Antonini*[2]

[1] Orange Labs
38 rue du G. Leclerc, 92794 Issy Les Moulineaux, France
{jeanmarc.thiesse, joelb.jung}@orange-ftgroup.com

[2] Laboratoire I3S, UNSA/CNRS
2000 route des Lucioles, 06903 Sophia Antipolis, France
am@i3s.unice.fr

ABSTRACT

2009 appears to be the launching date for several activities intented to challenge the video compression standard H.264/AVC. Improvements upon H.264/AVC can be achieved either by breakthrough or incremental approaches. Due to the extreme fine tuning of H.264/AVC, the former seems tough. The latter however seems more achievable, even with the announced 50% coding efficiency improvement objective.

In this highly challenging context, we propose a new scheme for macroblock partitioning and prediction. We extend the previously proposed Intra 1D scheme with a Skip 1D mode. Therefore, a macroblock is split in 1D partitions which can be predicted through two submodes in competition: Intra 1D or Skip 1D. As a second step, we further extend this scheme with an Inter 1D submode and provide the first keys to make this mode efficient. With the Intra 1D+Skip 1D scheme, an average 5.7% gain is reported compared to H.264/AVC, up to 10.1% for a given sequence. Preliminary results are also reported when the Inter 1D mode is added.

1. INTRODUCTION

Significant compression gain, compared to former video coding standards, has been achieved by the ITU-T SG16-Q6 H.264 standard of the Video Coding Experts Group (VCEG), also known as ISO/IEC JTC 1/SC 29/WG 11 MPEG-4 AVC [1] of the Moving Pictures Experts Group (MPEG). This gain results from the improvement of existing tools and the inclusion of new ones. These improvements concern mainly the motion estimation and the information coding with Context Adaptive Binary Arithmetic Coding (CABAC), but also the use of Intra and Inter modes with variable block sizes from 16×16 to 4×4.

Today, both VCEG and MPEG are working towards the improvement of H.264/AVC with projects respectively called NGVC and HVC. The goal is to reach a video coding standard that achieves 50% coding efficiency improvement, with a complexity increased by a factor 2 or 3. It can be done either by breakthrough approaches or by incremental approaches. This paper is classified in the latter category.

As a frame needs competition between spatial and temporal prediction to exploit differents image features, it is also the case at macroblock level. We propose a study for adding a new prediction mode, the Hybrid 1D. Built on the Intra 1D scheme [2], this mode makes a competition between three submodes (Intra, Inter, Skip) into a macroblock split in 1D partitions.

The remaining of this paper is organized as follows: a summary of H.264/AVC Intra and Inter prediction is presented in Section 2. The proposed Hybrid 1D prediction and coding is described in Section 3. Finally the experimental results are reported and analyzed in section 4.

2. STATE OF THE ART

2.1 H.264/AVC prediction and coding

H.264/AVC is an hybrid video coder: Intra and Inter prediction are used together to exploit respectively spatial and temporal redundancies. In H.264/AVC, this competition between Inter and Intra is performed at the macroblock level only.

Intra prediction is used to exploit the spatial correlation by using neighboring reconstructed pixels (the reference signal) to predict the current block. This is performed by three macroblock partitioning: 16×16, 8×8 and 4×4 with respectively 4, 9 and 9 predictors.

Inter prediction searches the best prediction for the current block in the reference frames by using a motion search algorithm. As successive frames are often highly correlated, Inter prediction generates a good prediction especially for video sequence with smooth motion. 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 block partitions are defined.

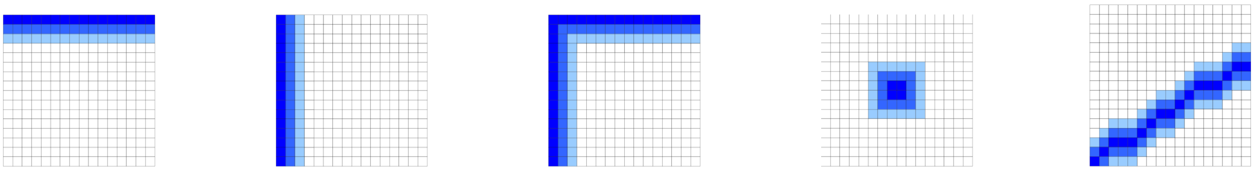
In addition, H.264/AVC has a Skip mode which predicts the current macroblock by the macroblock pointed in the reference frame by a motion vector predictor. This mode doesn't need any coding information besides his index signalisation and is often used in macroblocks without motion.

The choice to encode a macroblock in Inter, Intra, or Skip, is non normative. The JSVM reference software applies the following strategy: for each Intra and Inter mode (except Skip), the best predictor is selected and the difference between the selected predictor and the current block is transformed with a 4×4 or 8×8 floating point DCT and quantized. For each macroblock, the best mode is selected according to a rate-distortion (RD) criterion:

$$J = D + \lambda R, \tag{1}$$

where J is the RD cost, D is the distortion, R is the rate generated by the current estimated coding, and λ is the Lagrangian multiplier which depends on the Quantization Parameter (QP). The distortion D is computed by a SSE (Sum of Squared Errors) between original and predicted macroblock, the rate R includes the cost of the signaling of the mode and submodes for Intra and Inter 8×8, but also the transformed coefficients coding cost and the motion information (vector and reference frame) coding cost for Inter modes.

H.264/AVC has three different slice coding types: intra slice (I), predictive slice (P) and bi-predictive slice (B). For P and B slices, it is possible to have Intra coded macroblock but a given macroblock is always either I, or P or B.



(a) *Line* (b) *Column* (c) *'L'* (d) *Point* (e) *Edge*

Figure 1: Example of dilatation for the initial partition p_1 *Line*, *Column*, *'L'*, *Point* and *Edge*.

2.2 Recent contributions

Several modifications of the H.264/AVC modes have been recently proposed. An important part of these contributions concerns Intra coding by modifying the partitioning, scan orders or predictors. In [2], Laroche and al. propose a modification of all these attributes and the transform. To reduce the distance between the pixel to encode and its predictors, they propose to use 1D partitions with three scan orders, adapted predictors and a 1D-DCT.

Today, due to the emergence of HD 2K, 4K resolutions, Inter contributions are concerned with large partitions and transforms of size 32×32 and 64×64 [3]. Indeed, common partitions sizes from 4×4 to 16×16 have been well optimized during last years and studies about new partition is an interesting way to improve H.264/AVC coding. Thus, several works, as [4] [5], propose new geometry adaptive partitioning which up to now suffer from the cost of signaling the partitioning itself.

However, only a few works try to combine Intra and Inter prediction. In [6], authors propose to exploit both the correlations in temporal and spatial domains. They propose a new mode which is a weigthed combination of Inter 16×16 and Intra 16×16 Horizontal. The results, specially for high definition video coding, show that Inter and Intra combination is an interesting research topic because energies in the Inter and Intra residual signals are distributed differently and can be complementary. In [7], authors propose the *Second Order Prediction* which apply H.264/AVC Intra prediction to Inter prediction residuals. Although coding complexity increases because of the two prediction steps, the gains confirm that using Intra and Inter to predict a macroblock is efficient.

3. HYBRID 1D

The Intra 1D framework proposed in [2] has shown interesting results, consequently we propose to extend this method for Inter prediction with an Hybrid 1D mode. This mode differs from the recent state of the art by the fact that the cost of the signaling of the mode partitioning remains low, and that it mixes temporal and spatial prediction inside a single macroblock. Therefore, each 1D partition of a macroblock can be coded through Intra 1D, Skip 1D or Inter 1D submode in competition.

3.1 Intra 1D + Skip 1D

3.1.1 1D macroblock partitioning

The proposed 1D partitioning is similar to the one proposed in [2]. The main principle is reminded below.

The macroblock 1D partitioning is completely defined by an initial partition p_1 which is transmitted to the decoder as an index for the Hybrid 1D mode. The P other macroblock partitions, $p_i, \forall i = 1..P$, are obtained by propagation of the initial pattern p_1 thanks to a mathematical morphology dilatation operation using a 3×3 cross shape structuring element.

Examples of initial partitions and corresponding propagations are shown in Figure 1 for the partitions *Line*, *Column*, *'L'*, *Point* and *Edge*.

3.1.2 Scan orders

Three scan orders are proposed in [2] : the raster scan, the bi-directional scan and the hierarchical scan. We have yet implemented the first scan which consists on encoding partitions in lexicographic order 1, 2, 3, 4, etc... The others scan are more complex and permit to have several partitions as reference signal for Intra 1D.

3.1.3 Prediction

For each 1D partition of a macroblock, two submodes are in competition : Intra 1D and Skip 1D. The best predictor is selected for each submode and then the best submode is chosen through a RD criterion described in 3.4.

1. **Intra 1D**
   The proposed 1D partition takes benefit from the definition of specific predictors. In this section, the defined predictors are related to the *Line* macroblock partitioning and the raster scan order. These predictors can easily be generalized to the other 1D macroblock partitioning and scan order.
   In this case, for a partition p_i, the reference signal is the partition p_{i-1}. This is interesting because each pixel of the current partition p_i has a neighbouring pixel in the reference signal. Naturally the previous line is included as a predictor, and to exploit the spatial redundancies on the left of the current line, the left pixel and its two neighboring pixels on top and bottom are used. Other predictors are weighted means and shifts or combination of these predictors.
   As the partition p_{i-1} can be predicted by one of the two submodes, it is obvious that if p_{i-1} uses Skip 1D, p_i will be less efficiently predicted in Intra 1D because the reference signal will be worst. However, the left reference signal, the partition p_{-1} and all the Intra 1D or Inter 1D predicted partitions permit to conserve a satisfactory reference signal. Furthermore, as P and B frames take benefits of previous Intra frame, the macroblock encoding will take benefit of an Intra 1D predicted partition.

2. **Skip 1D**
   The second submode is Skip 1D. It is a particular case of Inter 1D. In this submode, a temporal prediction without information transmitted is done. A flag 1D is sent to sig-

nal that it is a skipped partition. The decoded partition corresponds to the partition predictor from the reference frame motion compensated by the motion vector predictor. The motion vector predictor is the one described for Inter 1D prediction in 3.2.

3.1.4 DCT 1D Transform

The standard H.264/AVC uses a 4×4 DCT and a 8×8 DCT for the residual coding of the Intra or Inter prediction. For the proposed scheme, we add a floating point 1D-DCT to fit with the linear shape of the proposed partitions. Concerning the quantization, we use the quantization defined for the 4×4 DCT. As Skip 1D partitions doesn't encode residuals, only residuals of Intra 1D coded partitions are transformed.

3.2 Inter 1D

As it will be reported in Section 4, the combination of the Skip 1D to the Intra 1D brings coding efficiency improvements. As another step forward, we have consequently extended the method by adding Inter 1D to the Skip 1D+Intra 1D scheme. In Inter 1D prediction, a motion estimation is done for each partition p_i and provides a motion vector $mv_i(mv_{i_x}, mv_{i_y})$. Pixels residuals (ε_i) and motion vector residuals (ε_{mv_i}) are transmitted. ε_i is obtained by substration of the current partition p_i and the pixels of the mv_i motion compensated predictor partition in the reference frame p'_i. ε_{mv_i} is obtained by substration of the motion vector mv_i and the motion vector predictor mv_pred_i.

$$\varepsilon_i = p_i - p'_i(x - mv_{i_x}, y - mv_{i_y}) \tag{2}$$

$$\varepsilon_{mv_i} = mv_i - mv\_pred_i \tag{3}$$

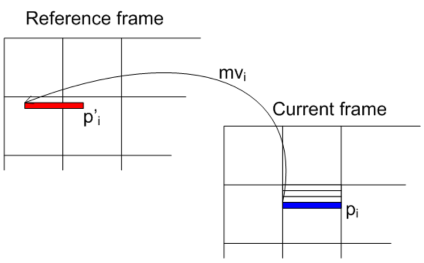


Figure 2: 1D partition matching.

The motion estimation uses a classical block matching algorithm adapted to the new partition size (Figure 2). In the case of the *Line* macroblock partitioning the full search attempt to find the best matching 16x1 pixels block. Half-Pel and Quarter-Pel are implemented too, using the standard H.264/AVC filters. For the initialisation of the motion search we use the motion vector predictor which calculation is described below.

As Inter 4×4, this new submode permits to have a better prediction of the partitions because there are less pixels to match. Consequently the distortion between the original and predicted partitions is reduced. As one mv_i by Inter 1D predicted partition has to be transmited, this submode is yet penalized by motion vector coding cost, this is why the motion vector prediction is important. We have implemented two motion vector predictions:

- the *first prediction* is a spatial median of the left, above and right-above macroblocks motion vectors. In this implementation, all the partitions p_i have the same motion vector predictor.
- the *second prediction* is the motion vector of the last previous partition coded in Inter 1D in the current macroblock. If there is no previous partition coded in Inter 1D, the *first prediction* is used.

3.3 Mode signaling

For signaling Hybrid 1D mode and submodes we use a tree structure which is depicted in Figure 3. The Hybrid 1D signaling is at the same level of the other H.264/AVC modes. Then a flag indicates the partition type (*Line*, *Column*, *Point*,...). Finally a flag for the *P* macroblock partitions indicates the partition submode (Intra1 D, Skip 1D or Inter 1D).

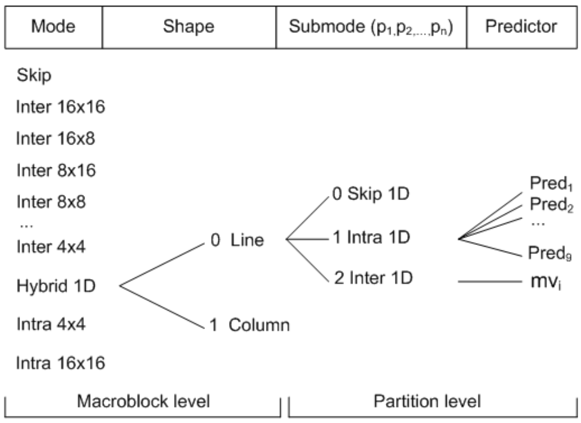


Figure 3: Hybrid 1D signaling.

Due to the large number of submodes contained in the Hybrid 1D mode, the signaling is a key issue. A particular effort has been made to efficiently encode the submodes. Built on the H.264/AVC Intra prediction submodes coding, we use a *most probable mode*. This tool permits to decrease the submode index coding cost by signaling if the current partition is coded through the same submode as the previous partition. Moreover, for the Intra 1D+Skip 1D scheme we choose to use an additional *shortcut* flag for signaling if all the remaining partitions have the same submode as the current one. If the *n* last 1D partitions of a macroblock are coded with the same submode, only the $P - n$ firsts flags and one additional *shortcut* flag are coded instead of *P* flags.

3.4 Competition criterion

The new Hybrid 1D mode competes with the existing Intra and Inter modes for each macroblocks. The selection of the best coding mode is made with the criterion (1).

For Hybrid 1D, the distortion *D* is computed by a SAD (Sum of Absolut Differences) between original and predicted macroblock after residual transformation using the DCT 1D for all non skipped partitions. *R* includes the cost of the mode signaling R_{mode}, the transformed coefficients coding cost R_{coeff}, the Intra predictor coding cost R_{Intra_i} if it is an Intra 1D partition and the motion information coding cost R_{Inter_i} if it is an Inter 1D partition (for Skip 1D partition, the cost is null) :

$$R = R_{mode} + R_{coeff} + \sum_{i=1}^{P} R_{Intra_i} + \sum_{i=1}^{P} R_{Inter_i} \quad (4)$$

As *Line* and *Column* partitioning have the same elements amount as the Inter 4×4, we use classical CBP with correspondence between partition p_i and 4×4 block number i.

Finally, this cost is compared to the cost of the other H.264/AVC modes and the coding which leads the lowest J is selected.

An additional criterion $J' = D' + \lambda R'$ is proposed for submode selection for each partition. For this criterion, D' is the luma distortion between original partition and reconstructed partition calculated with an SSE (Sum of Squared Errors) except for Skip 1D for which the DCT is not applied. The rate R', in addition of a macroblock mode signaling cost, is composed of:

- in Intra 1D: the sum of residual transformed coefficients cost, the signaling of submode and the current partition flag coding cost.
- in Inter 1D: the sum of residual transformed coefficients cost, the motion vector coding cost and the current partition flag coding cost.
- in Skip 1D: the current partition flag coding cost.

4. EXPERIMENTAL RESULTS

4.1 Experimental settings

The Hybrid 1D mode has been implemented into the H.264/AVC JSVM software (baselayer part only, scalable features of the JSVM are disabled). H.264/AVC reference results are generated using the Intra 4×4, Intra16×16 and all Inter modes, with the 4×4 DCT transform enabled. CABAC entropy coding method, known to provide the best results, is selected. The number of reference frames is set to 1. Experimental configuration is with a GOP (Group Of Picture) of 12 pictures: one I slice followed by eleven P slices are encoded, and P slices may contain Intra macroblocks. This GOP size corresponds to a typical configuration for many broadcast applications, where fast random access is required. Only the luminance component is processed (due to implementation issues).

Experiments have been performed on 5 well known sequences, with resolutions from CIF to 720p. 4 operating points, with QP equal to 22-27-32-37 are computed, providing realistic PSNR comprised between 30dB and 40dB. All bitrate savings are computed with the Bjontegaard metric [8] as recommended in the VCEG common conditions [9].

Results have been generated in 3 steps: Intra 1D alone, Intra 1D + Skip 1D secondly, then, addition of the Inter 1D. For these schemes, two initial partitions have been implemented: *Line* and *Column*. The scan order used is the raster scan and Intra 1D predictors are those defined in [2].

4.2 Bitrate savings

Table 1 gives the percentage bitrate savings for each sequence and for the three methods. For all sequences, the proposed schemes give a systematic bitrate saving. The average gains are 5.2%, 5.7% and 5.7% respectively in Intra 1D, Intra 1D+Skip 1D and Hybrid 1D schemes. The best result is obtained for Schumacher CIF which reaches 10.1% bitrates savings.

In average, the scheme with Intra 1D and Skip 1D gives a gain of 0.5% compared to only Intra 1D scheme and up to 1% for Spincalendar 720P. Hybrid 1D gains is approximately the same as the gain of Intra 1D+Skip 1D. At this stage we manage to compensate the cost of the signaling of the mode, and the cost of the mv_i thanks to a more efficient prediction.

Sequence	Intra 1D	Intra1D+Skip1D	Hybrid 1D
Schumacher CIF	9.8	10.1	9.9
City VGA	1.8	2.3	2.4
Exit VGA	2.8	2.9	2.9
Panslow VGA	8.8	9.2	9.5
Spincalendar 720P	2.9	3.9	3.7
Average	**5.2**	**5.7**	**5.7**

Table 1: Percentage bitrate savings of sequences with GOP 12 configuration for QPs 22-37 for the three differents methods *Intra 1D*, *Intra 1D+Skip 1D* and *Hybrid 1D*.

4.3 Results Analysis

	QP				
Mode	37	32	27	22	Average
Intra 16×16	4.8	4.0	3.2	2.7	3.7
Intra 4×4	2.1	3.0	4.0	6.1	3.8
Skip	73.7	66.2	54.9	32.5	56.8
Inter 16×16	12.3	14.8	17.7	26.0	17.7
Inter 16×8	2.0	3.4	5.8	8.8	5.0
Inter 8×16	1.8	3.4	5.4	7.8	4.6
Inter 8×8	0.4	1.4	3.9	9.5	3.8
Hybrid 1D	**2.9**	**3.8**	**5.1**	**6.5**	**4.6**

Table 2: Percentage of the selection of the H.264/AVC modes and Hybrid 1D for 4 QP on the 5 sequences tested.

Table 2 gives the percentage of selection of each H.264/AVC mode and Hybrid 1D mode. In average, for all QPs, Hybrid 1D has a percentage selection of 4.6%, it is higher than the selection of Intra 16×16, Intra 4×4 and Inter 8×8 and equivalent to the selection of Inter 16×8 and Inter 8×16. The Hybrid 1D selection increases from QP 37 to 22. It is explained by the mode selection performed by the optimal rate distorsion criterion which promotes selection of low cost modes like Skip or Inter 16×16 at low bitrate.

An example of the Hybrid 1D mode spatial selection for City VGA sequence at QP 27 is depicted on Figure 4 with *Line* and *Column* partitioning. We can see that the *Column* partitioning has more selections where the texture orientation is vertical which is the case for buildings, opposite of the Hybrid 1D *Line* which is selected on the top of the sequence where the texture orientation is horizontal.

Table 3 gives the percentages selection of each submodes. In average, for all QPs, the Skip 1D has a higher percentage selection than the others submodes. For the first scheme, Intra 1D+Skip 1D, Intra 1D has in average 22% of selection which allows to this scheme to improve coding because of the efficient Intra 1D prediction combined to a low cost Skip 1D submode. When Inter 1D is added, the Skip

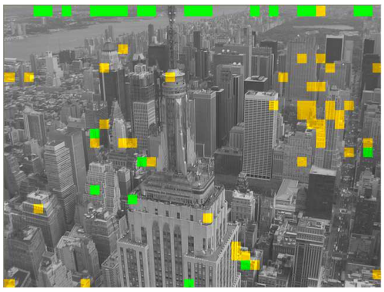


Figure 4: Spatial selection of the Hybrid 1D mode *Line* (green) and *Column* (orange).

1D selection doesn't change and the remainder selection is divided through Intra 1D and Inter 1D with a higher selection for Intra 1D. This confirms that the Inter 1D scheme still needs improvements. We can also notice that Inter 1D and Intra 1D have higher selection at high bitrate.

	Intra1D+Skip1D		Hybrid 1D		
QP	Intra 1D	Skip 1D	Intra 1D	Skip 1D	Inter 1D
22	24.8	75.2	16	74.3	9.7
27	22.2	77.8	15.2	77.2	7.6
32	20.3	79.7	16.1	78	5.9
37	20.2	79.8	15.5	80.1	4.4
Average	**21.9**	**78.1**	**15.7**	**77.4**	**6.9**

Table 3: Percentage of the selection of each submodes for the *Intra 1D+Skip 1D* and *Hybrid 1D* schemes for the 4 QP tested on the 5 sequences.

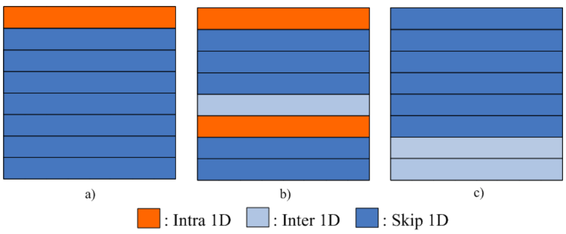


Figure 5: Submodes distribution examples in Hybrid 1D.

To conclude this analysis Section, we wish to highlight the rationale for providing such an Hybrid 1D mode. For that purpose, some typical submodes distributions are depicted in Figure 5. The first case 5(a) presents a macroblock with the first partition predicted in Intra 1D and all following partitions in Skip 1D, this case is interesting because the first partition is efficiently predicted with a close reference signal and this good prediction allows the remainder macroblock to be coded in Skip 1D with low cost. This type of macroblock coding can be more efficient than H.264/AVC Skip because a low coding cost is conserved with a better prediction for some 1D partitions predicted with Intra 1D.

The second case 5(b) represents the same configuration with in addition a partition which is predicted with Inter 1D and then the folowing partition can be better predicted in Intra 1D by using the efficient Inter 1D prediction as reference signal. It is typical when an edge cross the macroblock (not necessarily in the middle). In the last example 5(c), Skip 1D is efficient for the first lines, then, Inter 1D is used given that the motion vector derived fot the Skip 1D becomes less and less accurate.

5. CONCLUSION

In this paper, an extension of Intra 1D to an Hybrid 1D mode is presented. This mode proposes to code each 1D partition of a macroblock through Intra 1D or Skip 1D submodes. In a second step, an Inter 1D is added to this scheme in order to achieve a better prediction for 1D partitions. Efficient integration of Inter 1D requieres to solve some problems regarding the motion vector coding, the mode signaling and the submode selection. We propose technical solutions to these 3 problems and we report some preliminary results where the extra cost of 1D motion vectors is already compensated, compared to Intra 1D+Skip 1D.

In addition, the combination of the Intra 1D efficient prediction and the Skip 1D low coding cost inside a macroblock allows significant gains compared to H.264/AVC standard. The average bitrate savings reported are 5.7%. In the near future we expect to further improve this scheme by tackling the remaining difficulties related to Inter 1D.

REFERENCES

[1] "Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)," Standard Version 7: Apr. 2007, ITU-T and ISO/IEC JTC 1.

[2] G. Laroche, J. Jung and B. Pesquet, "Intra Prediction with 1D Macroblock Partitioning for Image and Video Coding," *VCIP'09*, San Jose, California, Jan. 2009.

[3] P. Chenn, Y. Ye and M. Karczewicz,"Video Coding Using Extended Block Sizes," ITU-T VCEG contribution C123, Geneve, Switzerland, Jan. 2009.

[4] O. D. Escoda, P. Yin, C. Dai and X. Li, "Geometry-adaptive block partitioning for video coding," *ICASSP 2007*, Honolulu, U.S.A, Apr. 2007.

[5] K. Vermeirsch, J. D. Cock, S. Notebaert, P. Lambert and R. Van de Walle, "Increased Flexibility in Inter Picture Partitioning," *MMSP 2008*, Cairns, Australia, Oct. 2008.

[6] J. Xin, K. N. Ngan and G. Zhu, "Combined Inter-Intra prediction for high definition video coding," *PCS 2007*, Lisboa, Portugal, Nov. 2007.

[7] S. Li, L. Yu, J. Wang and L. Xiong,"Test results of second order prediction (SOP)," ITU-T VCEG contribution C154, Geneve, Switzerland, Jan. 2009.

[8] J. Jung and S. Pateux, "An Excel add-in for computing Bjontegaard metric and its evolution," ITU-T VCEG contribution VCEG-AE07, Marrakech, Marocco, Jan. 2007.

[9] T. Tan, G. Sullivan and T. Wedi, "Recommended Simulation Common Conditions for Coding Efficiency Experiments," Input/Discussion VCEG-AE10, ITU-T VCEG, Nice, France, Oct. 2005.